Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.7, No.6, 2016
www.iiste.org

Assessment Skills: A Case of Mathematics Examination and Its Place in Math-Teacher Development

Ahmad Hassan Qudah
Al al-Bayt University, Mafrag - JORDAN.

Abstract
The research aims to reveal the specific way to evaluate learning mathematics, so that we get the " measuring tool" for the achievement of learners in mathematics that reflect their level of understanding by score (mark), which we trust it with high degree. The behavior of the learner can be measured by a professional way to build the test, through a sequence of steps that make the mark obtained has full confidence, and reflect the real level of the learner.

Introduction

Recent research in the field of education has been directed towards bridging the gab between theory and practice. Bassey (1995) defines theoretical researchers as one who try to describe, interpret, and explain events without making any judgment about them. Today, researchers attempt to describe, interpret, and explain events in order to change them for the better. Alshannag (2001) ,the aim of this kind of research is to encourage teacher to reflect on their teaching and become more aware of their strengths and weakness.

Therefore, the paper discuses the main phases of evaluation, which induces constructing, implementing, scoring and analyzing examinations.
This research aims to provide the reader with the skills to build and develop the exams grades in mathematics, so we are confident the results of this test. At the same time, It has the aim to answer the following questions:1) Is it possible to find a measurement tools that measure a student's behavior to be measure ? 2) Is there a way to build a professional exams grades? 3) Can you trust the results of the exams that are built in a professional manner?.
The researcher did not find so many studies on the limits of his knowledge and brief him on the quality of the studies on the construction of the exams in a systematic manner, which motivated him to walk in this direction.
Appropriate to use a method of qualitative research, such as in this case being the most appropriate in the case of building tests.

A very important question in teaching is: where is the student now? To find the answer to this question, a teacher gives an examination, which at the same time achieves some further objectives:

- Giving the student a positive feeling towards the subject matter,
- Building confidence in the student that he/she can do the work by his/her own activities,
- Encouraging a questioning attitude in response to a pupil's curiosity regarding subject-matter problems. (NCTM ,2000)

It is important to consider the objectives to each question in the test paper and even more important to relate those objectives to class activities and the type of homework done before the exam. Teachers should state their questions very clearly in a test paper, using terminology appropriate to the pupil's level of understanding.(Odeh,2005)

The training program for student teachers in universities is combination of content and mythology. Teachers focus their attention on content, lesson planning, questioning, home working, and testing, and must come proficient in these essential areas in order to give effective instruction. In the area of testing it is important to have knowledge of what to test as well as how to test.(MOE,2007)

In many cases the result of examinations can be utilized for different purposes: (1) for classifying student into instructional groups, (2) for assigning levels of performance mastery in assessment, (3) for diagnosing specific weakness.(Al Kelany,2005)

Exam results should be analyzed so that teachers are aware of the type of errors that have been made so that emphasis can be placed on the particulars in their future teaching. This information is used to guide the next steps in teaching.(Abu Zeneh, 1998).

The current study will focus on the most important details needed to build a math teacher math exam. The following steps show the methods:

1. Constricting Method:

Exams should be planned in advancing connection with planning the teaching process. It is even worth wile to make up the next exam as soon as the previous one is given and after the homework assignments have been planned.

The teacher should begin to construct the exam by listing the topics with respect to homework assigned and class work; he/she should also consider the objectives of each question.

The type of questions that can be asked is single-concept questions, multi-concepts questions, multiple choice questions, and essay questions.

The exam should begin with easy question and the proceed with progressively more difficult ones topics converted on last three days before the exams should not be included since students would need to practice these topics through homework first. Students should be informed as to the number of points per question on the test and the directions to be followed so that they can budget their time accordingly.(Abu Zeneh,1998)

Exam question papers should be neat and legible and the pages not crowded. This study suggest writing a draft of the actual first, and then making changes if necessary.

Finally, when the exam is constructed the following questions should be considered: (Abu Zeneh,2012)

- When should the exam be given?
- What is to be tested?
- What types of questions should be asked?
- How long should the exam be?
- How difficult should the exam be?
- What should be looked at while making the exam?
- How should the exam results be used?
- should the exam be reviewed in class after it is marked? When? How? How to future tasks relate to the exam results?

The following part of constructing exam consist of five major subtitle, including main ideas, objectives, content analyses, specification table, and writing question.

1.1 Main Ideas

The teacher should be able to identify the main ideas are included in a topic, and a student teacher should have read a great deal about the topic and have discussed any topics or activities unitized in the textbook before he/she classifies anything as a main idea.

In many cases the textbook itself is designed such that the topic consists of many headlines or tasks. In this case the role of the teacher should be to identify the main ideas as this is very important for testing and planning lessons in a subject area.

1.2 Objectives

The teacher should derive the main objectives of a topic just as he/she identified the main ideas, perhaps by referring to detailed lesson planes for each topic.

The objectives should be stated clearly and simply and written in a form such that the answers are measurable. Some objectives should ultimately be elicited from the pupil. It need not, however, be elicited at the very beginning of the period.(Odeh,2005).

Example of objectives:

- To factor a quadric trinomial.
- To compare tow rational numbers with unlike denominators.
- To find the product of two binomials.
- To solve problems involving areas of complex figures.
- To find the relationship.
- To solve new puzzles based on the geometry.

1.3 Content Analyses

The topic should consist of various elements such as concepts, facts, generalization, vocabulary, exploring, and others.

The student teacher should learn to categorize the knowledge into parts insofar as possible.

This means the content will be divided into small tasks or into small parts such that each one has a main feature.

For example: A mathematics lesson should consist of four main elements of mathematics: (Abu Zeneh,2012)

1. Concepts and symbols
2. Skills
3. Generalizations and formulas
4. Problem-solving approach

At this point it is important to remember that not even practice will make you a perfect teacher; there is no such person. What it will do, however, is make you an interested teacher, an inquiring teacher, and therefore a better teacher. The secret of your success will lie in own efforts toward self-improvement.

1.4 Specification table

A specification table is a major part of processing or constructing an exam. Such a table describes the content and confronts the topics of content with bloom's levels of performance for each question.

In table 1, vertically, each cell represents the content of the question and, horizontally expresses the levels of Bloom. The following table 1 illustrate the description of each cells according to levels and content.

Table 1: percentage of levels of bloom (example)

Levels of bloom **Content analysis**	**Gaining knowledge**	**Application**	**Analysis**	**Synthesis**	**Evaluation**	**Percent of each part of content analyses**
Concepts of fraction	**Q1**	**Q2**				**20%**
Addition on fraction		**Q3**		**Q4**		**25%**
Subtraction on fraction		**Q5**	**Q6**			**20%**
Division on fraction		**Q7**	**Q8**	**Q9**	**Q10**	**35%**
Percent of each level	**10%**	**40%**	**20%**	**20%**	**10%**	**100%**

Table 1 shows the percentage of each part of content analysis which, in the last column and also in the row, clearly gives an idea of percentage of levels of bloom. The teacher has to determine these percentages. The objectives and main ideas written in previous steps can be translated into the cells of the table which are denoted Q1, Q2, ... Q10.

1.5 Writing questions

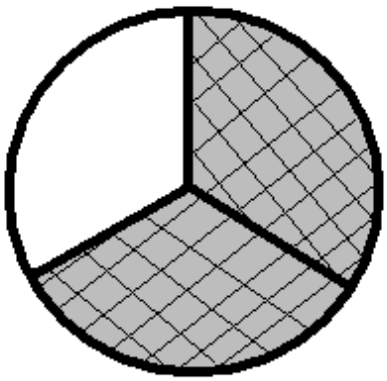
The step of writing questions involves a transformation from the specification table to the final form of the questions.(Abu Zeneh,2012)

It is as follows:

Q1: the content is: the concept of a fraction bloom's level is gaining knowledge.

So, Q1 will be in many cases (there is a margin of freedom for the teacher):

Case 1: express the shape of the region in the following figure using a fraction:



Case 2: suppose that you eat 5 sweets from a box which contains 18 sweets. What fraction describes the portion still in the box?
Case 3: design a representation (shape) for the fraction:
3/4, 1/5, 1/6

Others questions can be developed in the same way. Then the questions should be written in a form which is comfortable for this pupils.
You can change the order of the questions as follows:

Q1	Q2	Q6	Q4	Q10
	Q3	Q8	Q9	
	Q5			
	Q7			

This classification is derived from specification table such that the question progress from easy to difficulty and then to more difficult question.(Odeh,2005).

2. Holding the Examination

A test is conducted in an atmosphere which is relaxing for both the students and the teacher; therefore, no tension or stress can be detected during the test period.

It is preferable that the teacher remains at distance from the students during the test period so that other students and observers will not misunderstanding the movement.

Clear instruction must be written on the first page so that there is no need for questions during the test, this will also prevent the teacher from being required to answer any question, The test question must be explicitly specified in the appropriate position. (New York City Board of Education,2005).

The test papers must be distributed as quickly as possible so that no time is wasted. After the test period the test papers should be derived to the teacher all at once in a systematic manner.

3. Scoring Exams

When scoring exams, the teacher should mark one question on all papers and then proceed to the next question on each paper, etc. in this way the marking will be consistent in all student's test. If an answer is wrong, an "x" should be put through the answer so that it cannot be changed without being noticed.

Besides putting the grade on the test, whenever possible, a teacher should try to write a comment that will encourage better performance or give credit to the pupil's work. (Abu Zeneh,2012).

Try to have the test marked by the day after it is given so that the most can be gained in learning from mistakes, witch are still fresh in the pupil's minds. It is necessary to inform the pupils of the number who have failed. This type of discussion should be avoided since it is usually causes more problems rather than helping to improve instruction. (Odeh,2005)

If test papers are returned the very next day after the test has been given, it is still fresh in the pupil's minds. It also shows the pupils that the teacher is concerned about the improvement in learning in an individual subject.(Odeh,2005)

A discussion should take place with the pupils as to the purpose of going over a test, e.g., looking for points of error and looking for alternative methods.(Abu Zeneh,2012).

Ultimately, the pupils must feel that they have something to aim from this experience and the understanding what they got wrong has value for the future, especially in subjects which are learned in sequence.

4. Tabulating Marks

The teacher should grade each question individually because it shows the pupil's performance on each part of the exam. Furthermore, the teacher should focus on the response of the pupil on each part of the task, so that the mark will be good indicator of knowledge of the content of the task.(Byrnes,H. ,2008).

Formerly, teachers gave marks for the entire exam. However, current practice a partial mark is given for each question. The following table shows the distribution of the partial marks with regard to the final mark.

Table 2: distribution of partial marks with regards to the final mark

Name	Q1	Q2	Q3	Q4	Q5	Final exam
Ahmad	7	4	10	10	10	41
Ali	0	6	5	7	8	26
Ibraheem	1	0	7	6	6	20
Eman	9	7	8	9	19	42
Sanna	5	6	8	7	10	36
Mysoon	0	4	4	6	7	21
Ead	8	7	7	6	9	37
Kaled	4	6	7	5	5	27
Fareed	3	4	5	6	7	25
Salem	5	4	3	7	9	28
Samya	6	7	2	3	4	22

Note the situation of Ahmad, who got 41/50 points, which is very good, but the partial mark of question 2 reflects his weak performance here.

5. Calculating the quantity Indicators

The following part consist of five quantity indicators, including discrimination index, difficulty index, the mean, standard deviation, and consistency.

5.1 Discrimination index (di)

The discrimination index in an indication of how a given item can discriminate between the high-scoring and low-scoring groups, but it does not lend it self to providing a direct comparison of the degree of discriminating power of the individual items in a test. It fails in this respect because the high scoring and low-scoring groups.

In 1939, Kelley demonstrated that the lowest and highest 27% of a sample are optimal when item analysis data are to be obtained for items of 50% difficulty level and low reliability that are scored in graduated amounts. These groups are, as he pointed out, ordinarily the most practical for use in item analysis even though the items are scored only as "satisfactory" (passed) or "unsatisfactory" (failed).

If the discrimination index for a question equals zero or is a negative value, then a question does not discriminate between pupils, which is bad, and the question must be dropped or improved. We consider the following cases: (Miller,2004)

Case 1: if the discrimination index for a question shows an interval of 0 to 20%, then the discrimination index for a question is weak.

Case 2: if the discrimination index for a question shows an interval of 20% to 40%, then the discrimination index for question is intermediate (midpoint).

Case 3: if the discrimination index is more than 40%, then the discrimination index for the question is high, regardless.

In general, when the discrimination index approaches the high level, the question is good.(Odeh,2005).

5.2 Difficulty Index (pi)

The difficulty of an item can be defined as the proportion of sample of tested pupils who marked the item correctly an all tested pupils.

For example:

The following marks are taken from the results of a mathematics exam in the 9^{th} grade.

Table 3: example for the difficulty index (pi)

Name	Q1	Q2	Q3	Q4	Q5	Total mark
A	1	1	1	1	1	9/10
B	1	1	1	1	0	8.5
C	1	1	0	1	1	8
D	0	1	1	1	0	6
E	1	1	1	0	1	7.5
F	0	1	1	1	1	8
G	0	0	0	1	1	4.5
H	1	1	1	0	0	5
I	1	0	1	0	1	4.5
J	0	1	0	1	0	4
K	1	1	1	0	1	7.5
L	1	1	1	0	1	6.5

1: pupil's performance is satisfactory;
0: performance is unsatisfactory.
To find the difficulty index (pi) for 5 questions one would proceed as follows:
P1: we find 12 positions and 8 ones in the column of Q1, therefore we have:
P1=8/12=0.67 p2=10/12=0.83 p3=9/12=0.75 p4=7/12=0.58
p5=8/12=0.67

To find the discrimination index (di) for these questions, we apply the following:
(a) Record the table with respect to total marks from top to bottom:
(b) Compare 27% of the group (i.e.; 27% ×12=3.24: 3pupils).
(c) The highest 3 and the lowest 3 total marks of the group.

Table 4 illustrates this:

Table 4: example for the discrimination index (di)

Name	Q1	Q2	Q3	Q4	Q5	Total mark
A	1	1	1	1	1	9
B	1	1	1	1	0	8.5
C	1	1	0	1	1	8
. . .						
G	0	0	0	1	1	4.5
I	1	0	1	0	1	4.5
J	0	1	0	1	0	4

Now it is easy to compute the (di) as follows:
1st step: Divide the number of ones of A, B, C by 3 (highest group) and divide the number of ones of G, I by 3 (lowest group)
2nd step: subtract the fraction of lowest group from the fraction of highest group.
d1=3/3-1/3=2/3=0.67
d2=3/3-1/3=0.67
d3=2/3-1/3=1/3=0.33
d4=3/3-2/3=1/3=0.33
d5=2/3-2/3=0
The question is more difficult when the difficulty index approaches 0.
The question is easier when the difficulty index approaches to 1. The question is acceptable and very suitable when the difficulty index lies between 30% and 70%.(Odeh,2005).

5.3 The Mean

The most commonly used indicator of the central tendency is mean. The mean is nothing more than arithmetic average of scores in a distribution, calculated by adding the scores and dividing this number by

the number of cases in the distribution. Thus, the mean should give us an idea about the level of total group performance. At the same time we can use this indicator (mean) to compare the performance of an individual pupil, as follows:

:mean; xi: scores of the pupil $\overline{X}$

If xi = $\overline{X}$, the performance of pupil is equivalent to moderate.

If xi > $\overline{X}$, the higher the mean the better the pupil's performance.

If xi < $\overline{X}$, the lower the mean the poorer the pupil's performance.

5.4 Standard Deviation

Standard deviation is the square root of an average of the differences (deviations) between the scores of a distribution and its mean. The standard deviation computation is straightforward, with one twist. First, the mean of the distribution is calculated and then it is subtracted from the value of each score, thus yielding a series of deviation scores. To illustrate this: if the mean of distribution were 10, that number would be subtracted from all scores in the distribution, for example, 12-10=2;16-10=6. But, of course, there will be scores lower than the mean, for example, 8-10=-2; 7-10=-3. Subtracting the mean form these scores yields negative deviation scores.

If we were simply to add all these deviation scores, the result would be 0. So, before trying to compute an average of these deviation, we first square each deviation because that operation gets rid of minus signs (a negative number times itself yields a positive product: (-5) × (-5) = (+25).

These squared deviations are first added and then divided by the number of scores in the distribution, just as in computing the mean. Finally, to get back to the number size we were working with originally, we take the square root of the summed square deviations.

The computation procedure for the standard deviation SD can thus be represented as follows:

$$SD = \sqrt{\frac{\text{Sum of Squared Deviation Scores}}{\text{Number of Scores in the Distribution}}}$$

If you think for a moment about the procedure involved in computing the SD, you will realize that the more distant the scores of a distribution are from its mean (i.e., the farther they are spread out), the larger the standard deviation will be. Thus, in describing two sets of pupil scores on the same test, we indicated that the SD of group x was 10.4 and the SD of group z was 6.7. We would know that the scores in x were much more variable (i.e., spread out from the mean) than scores in group z. used in combination with the mean, SD is an extremely useful way of reflecting group variability and is undoubtedly the most widely employed index of variability that evaluators will encounter.

5.5 Consistency

Internal consistency is a method of estimating a test's reliability.(Lantolf,J.(2008).

Suppose we construct a test containing ten items and administer the test to 20 pupils. The number of items and objects taking are important in measuring consistency.

Kuder and Richardson developed two formulas for measuring consistency that can only be scored as right or wrong. There are two major considerations in the use of internal consistency estimates. First, this method should not be used for speed tests that are not completed by all those being tested. Second, it provides no estimate of stability over time.

The KU-RI21 (Kuder and Richardson-21) formula is some what less accurate than the KU-RI20 formula, but it is so simple to compute that it is probably the most frequently employed estimate of internal consistency.

$$KU-RI21 = \frac{K}{K-1}\left(1-\frac{M(K-M)}{KS^2}\right)$$

Where K is the number of items in the test, M is the mean of the set of test scores, and S the standard deviation of the set of test scores.

$$KU - RI20 = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} Pi(1-Pi)}{S^2}\right)$$

Where K: number of items pi: difficulty index
S: variance of the set of test scores

Like other internal consistency approaches to reliability, the (Kude_Richardson) method focuses on the degree to which the items in the test function in homogeneous fashion.

Thus, the coefficient will be larger (closer to 1.00) when the test items are Interco related. This is the meaning of internal consistency reliability.

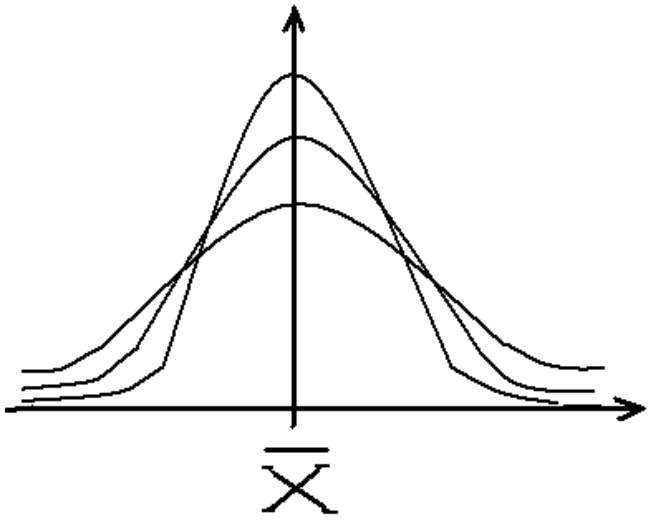
6. Graphic Representation

The mean is arithmetic average of the scores and represents the balance point of the distribution. The standard deviation describes the spread or clustering of scores in distribution.

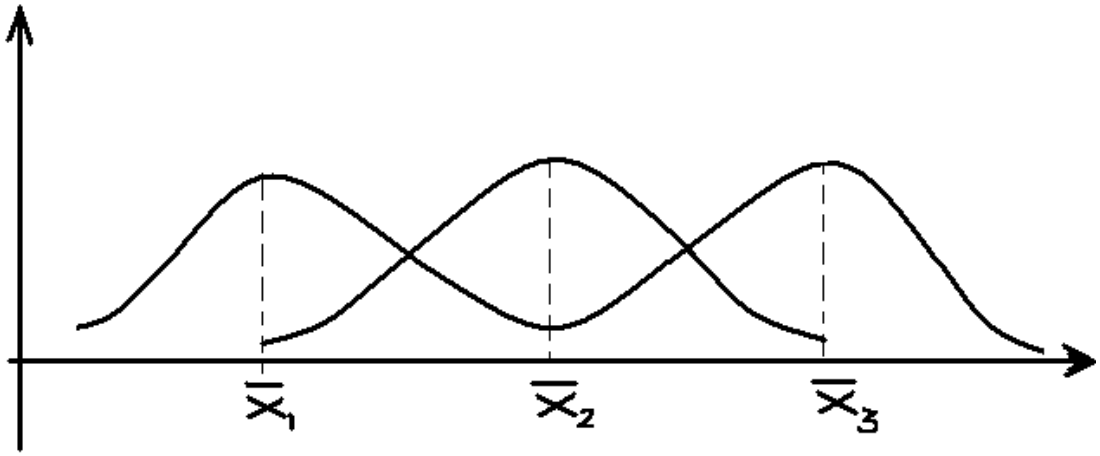
Interestingly, there are only two things we need to know about the normal curve in order to draw it: its mean and its standard deviation. In other words, the equation giving the high of the curve (y) for any particular point on the horizontal axis (x) can be solved, provided we know the values of the mean and the standard deviation.

The partial significance of this point is that once we know that a variable is distributed normally, with a particular mean and standard deviation, we have a perfect description of the entire distribution. We can, in effect, specify precisely what proportion of the observation will fall between any two values we care to specify.

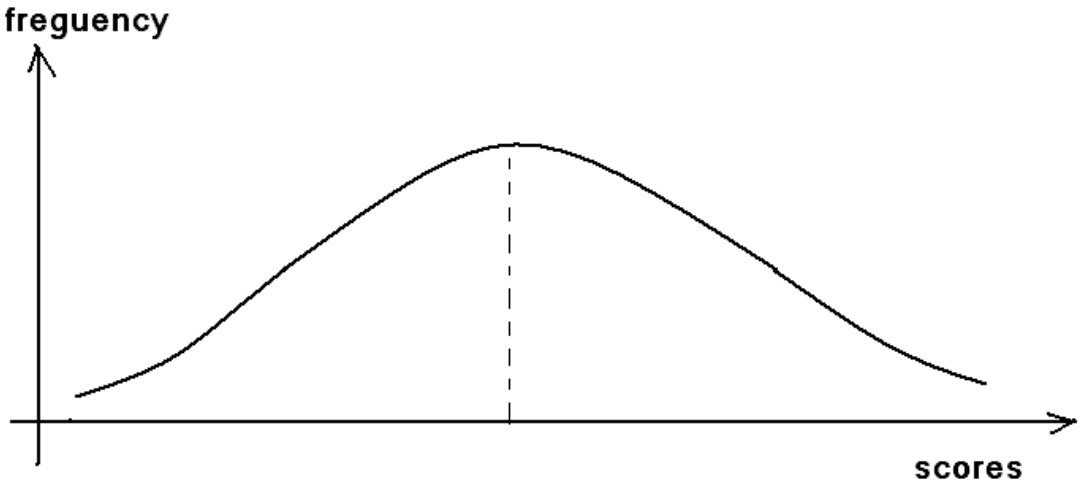
To be more specific, all normal distribution have the following property: if we draw a vertical line to indicate a fixed number of standard deviations above the mean, we cut off a constant proportion of distribution of scores. For any number of standard deviation above the mean it is possible to specify what proportion of scores falls above this value and what proportion falls below it (since the total area covers 100 percent of the scores). Furthermore, since the curve is symmetrical, a line draw a given distance below the mean will cut off the same proportion of scores as the one drawn that distance above the mean.(Abu Zeneh,2012). The following distributions illustrate the situations of tests in general without focusing on the items of a test.



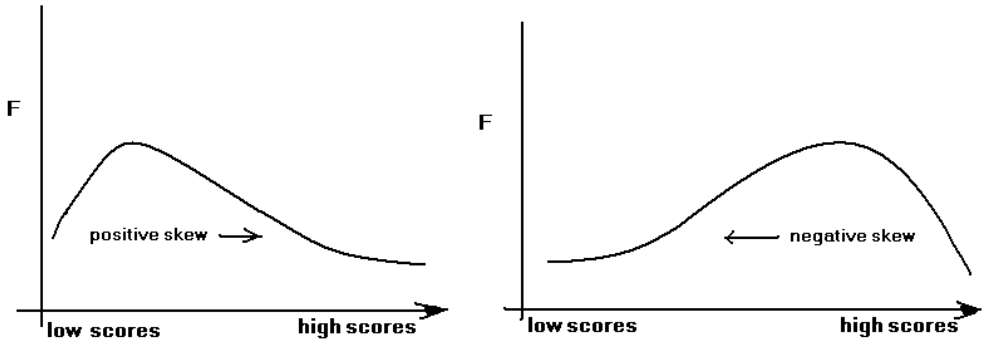
One mean $\overline{X}$ of 3 different standard deviations



One standard deviation of 3 different means



Distribution of scores for the very best test



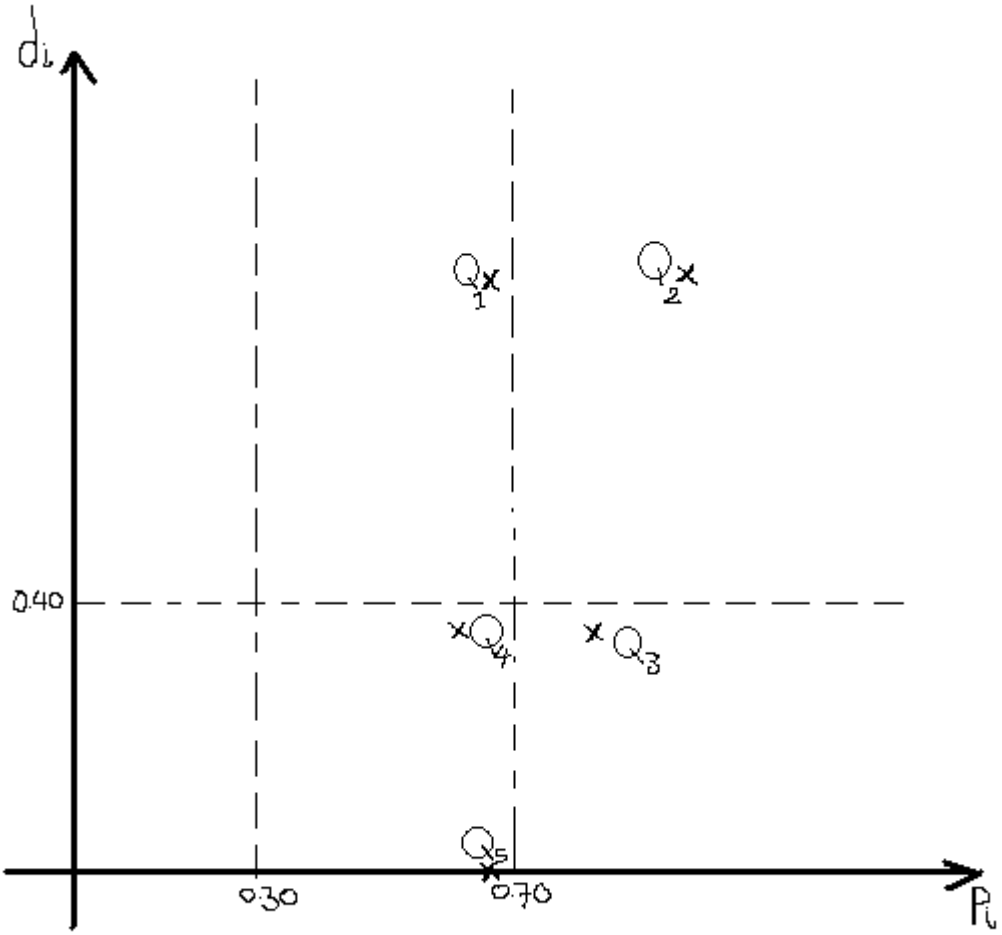
Distribution of scores on hard test distribution of scores on an easy test

The following table (table5) illustrates the items-analysis data for the exam presented in table3:

table 5 Difficulty index & Discrimination index for five questions

Questions	Difficulty index	Discrimination index
Q1	**0.67**	**0.67**
Q2	**0.83**	**0.67**
Q3	**0.75**	**0.33**
Q4	**0.58**	**0.33**
Q5	**0.67**	**0**

We can represent these data graphically as follows:



A best question that: 0.03 ≤ pi ≤0.7 and di ≥ 0.4
An excellent question is shown by: 0.3 < pi < 0.7 and di > 0.4.

For locating the region of accepted or rejected question, we consider the previous graphic, and focus on two consideration: first, the intervals of accepted difficulty index between 0.3 and 0.7 ; second, the accepted level of discrimination index, there is no cut-off score or level but it is better to be over 0.40. Accordingly, we can only accept question number 1 and must reject the others.(Davison,C.2007).

In many cases we need questions such as demonstrated by question number 2, which discriminates among students and seems easy.

7.Results:

The exam constructed as follows:

- Identifying main ideas,
- Deriving the objectives,
- Preparing content analyses,

- Preparing specification table,
- Writing questions.

Holding exam in an atmosphere which is relaxing for all.
Correcting questions one by one from all papers.
Tabulating the marks by each question separately.
Calculating the quantitative indicators.
Graphic representation of the indicators.

8.Discussion:

The questions of this study focused on preparing math-examination & analyzing the results, in order to give a credit evaluation for the students achievement.

This section is about design- making among the quantity indicators that computed about the exam, It is divided into four major subtitle, they are modifying the mark, horizontal treatment, vertical treatment, and percentage of the problem for each task.

8-1 Modifying The Marks

If questions are rejected on the above graphic, then the teacher should subtract the marks of rejected questions from the total marks so that the new total marks appear. The new marks are more valid the previous mark. The following example illustrates this procedure. If Q3 is rejected:

Before modifying	
Q1	**10/10**
Q2	**10/10**
Q3	**5/10**
Q4	**7/10**
Total	**32/40**

After modifying	
Q1	**10**
Q2	**10**
Q3	**R**
Q4	**7**
Total	**27/30**

You can see that 27/30 better reflects the pupil's performance on the exam than 32/40, because question 3 (rejected) is not suitable for statistical evaluation.

8-2 Describing the pupils Performance (Horizontal Treatment)

After modifying the marks, we can consider the matrix of student's achievement. All these processes depend on the indictors and their graphic representations. At the end of the processes we have a collection of marks and names which fully signify pupil's performance. We are then able to focus on each case and describe the weak and strange performance for each individual pupil. Table 6 shows a matrix which gives the full description of each pupil for each task:

Table 6 : Achievement matrix of pupils

Name	**Q1 Out of 5**	**Q2 Out of 10**	**Q3 Out of 4**	**Q4 Out of 9**	**Q5 Out of 7**	**Total marks out of 35**
A	**3**	**9**	**4**	**6**	**7**	**29**
B	**4**	**4**	**4**	**4**	**4**	**20**
C	**5**	**5**	**4**	**8**	**6**	**28**

Consider pupil B: his performance is below the usual standard in tasks Q2, Q4 and Q5 but is satisfactory in others.

The teacher could be more specific when describing the content and level of each task for witch B has performed poorly.

8-3 Describing the pupils Total Achievement (Vertical Treatment)

The total achievement of pupils is very clear if viewed vertically to the achievement matrix. If you focus on the first two items and compare the total marks that the pupils received, the result 12 and 18, respectively.

So, 12/15=0.80 represents very good achievement for the class and 18/30=0.60 weak achievement for the class.

If you calculate this for the 3rd question, you get 12/12=11.00, which represents excellent achievement for the class.

Try to compare the total achievement for the 4th and 5th question in the same way.

8-4 Calculating the Percentage Of the Problem for Each Task

The percentage of difficulty for any question according to the total achievement for the same question by taking complement can be calculated as follows:

Suppose that the total achievement of question is 0.64. This implies that the percentage of the difficulty is 03.6 (0.46+0.36=1.00).

If we have five tasks and their achievement is represented by 0.37, 0.60, 0.65, 0.77, 0.85, and then their percentage difficulty size are 0.27, 0.40, 0.35, 0.23, and 0.15, respectively. Therefore, the teacher should solve the problems by considering the percentage difficulty sizes. The order will be 0.40, 0.35, 0.27, 0.23, and 0.15, from a difficult task to the easier one that appears in the 5th task.

This sequence of percentage should show the teacher how to deal with the content of the questions that were given in exam. It is important to start from the content of the more difficult task and proceed to the content of less difficult ones.

In fact, the teacher should refer to specification table to note the content and level of each task that was presented as problem. He/she should take this into consideration when designing future tasks related to these contents.

Conclusion: At the end of the research, the researcher draw the final conclusion by depending on the results as follow:

The behavior of the learner can be measured by a professional way to build the test, through a sequence of steps that make the mark obtained has full confidence, and reflect the real level of the learner.

References

Alaam, salah aldeen (2001). Diagnostic tests reference stake in educational and psychological training areas, Cairo, Dar Alfeker.

Boyles, P. (2005). Assessment literacy. In M. National Assessment Summit Papers. pp.(1 -16), http://www.nflrc.iastate.edu/newsite.

Byrnes, H. (2008). Assessing content and language. language testing and assessment, Vol.(7), pp(37-55).

Davison, C.(2007). Views from the chalk face: Language Assessment Quarterly, (4), (1), p34-68.

Falsgraf, C. (2005). National Assessment Summit. Meeting conducted in Alexanderia, Va.

Kunnan, A.(2004). Regarding language assessment. Language Assessment Quarterly, (1), p,1-20.

Lantolf, J.(2008). Dynamic assessment. New York: Springer Science and Business Media.

Leung, C.(2004). Developing formative teacher assessment: Knowledge practice & change. Language Assessment Quarterly, (1),p.19-41.

Lynch, B& Shaw, P.(2005).Portfolios, Power and Ethics, TESOL Quarterly, (93),(2),p.263-298.

Odeh, Ahmad (2005). Evaluation and Assessment in the Educational Process.3red Edition, : Dar Al – Ammal , Irbid - Jordan.

MOE (2007).Assessment Strategy and its Tools (Theoretical Frame) .Editions of Ministry of Education.(MOE).

MOE (2008). Teacher's guide in the construction of tests.

Malone, M. (2008). Training in language assessment. New York: Springer Science and Business Media.

McMillan, J.H.(2000). Fundamental assessment principles for teachers and school administrators. Practical Assessment Research and Evaluation, (7), (8), p.123-147.

City boared of education (2005). Staff development for new teacher mathematics: High School New York

Davis, Fredrick B. item-analysis Data (Harvard Education Papers,Number 2).

Miller, Steve (2004) Experiment Design and Statics, 2nd Edition.

Popham, W. James (1993). Education Evaluation.

Porkess, Rodger (1982). Collins Educational; Mathematics; A Complete Course For First Examination.

Richard & Clifford (1997). Mathematical Question from the Classroom.

Swalvia & Yesseldyke (1981) . Assessment in Special and Remedial Education.

Bassey , M. (1995). Creating Education Through Research . Whitechead (Eds.). London & New York :Hyde Publication.

Alshannag, Q. (2001).Conducting Classroom Research to Improve Teaching : Enjoying Teaching : A Handbook of Practice- orinted Teacher Education, Editions Mathieu, Heidelberg.

Stiggins, R. (2001). Secondary teachers classroom assessment and grading practices. Educational Measurement: Issues and Practice, 20,(1), 20-32.

Stoynoff, S. &Chapelle, C. (2005). ESOL tests and testing: A resource for teachers and program administration. Alexandria, VA: TESOL publications.

Winndschitl, M.(2002). Framing constructivism in practice: An analysis of the conceptual, pedagogical culture and political challenges facing teachers. Review of Educational Research,(72),(2),p. 131-175.